Spiess, Hamerle:

On the properties of GEE estimators in the presence of invariant covariates

Sonderforschungsbereich 386, Paper 13 (1996)

Online unter: http://epub.ub.uni-muenchen.de/

On the properties of GEE estimators in the presence of invariant covariates

Martin Spiess and Alfred Hamerle*

Abstract

In this paper it is shown that the use of non-singular block invariant matrices of covariates leads to 'generalized estimating equations' estimators (GEE estimators; Liang, K.-Y. & Zeger, S. (1986). *Biometrika, 73(1)*, 13–22) which are identical regardless of the 'working' correlation matrix used. Moreover, they are efficient (McCullagh, P. (1983). *The Annals of Statistics, 11(1)*, 59–67). If on the other hand only time invariant covariates are used the efficiency gain in choosing the 'correct' vs. an 'incorrect' correlation structure is shown to be negligible. The results of a simple simulation study suggest that although different GEE estimators are no more identical and are no more as efficient as an ML estimator, the differences are still negligible if both time and block invariant covariates are present.

Key words: Generalized estimating equations; Invariant covariates; Asymptotic properties.

1 Introduction

The 'generalized estimating equations' approach (GEE approach) proposed by Liang and Zeger (1986) and Zeger and Liang (1986) has been the subject of many papers, some of them investigating the properties of the corresponding GEE estimators (e.g. Fitzmaurice, Laird, and Rotnitzky, 1993; Sharples and Breslow, 1992; Hamerle and Nagl, 1987). Although in most regression models for correlated responses used in practical applications different kinds of covariates are included, e.g. block invariant covariates modeling time or within-factor level effects or time invariant covariates modeling between-factor levels or sex of the observed objects in the sample, their effect with respect to the properties of the GEE estimators have routinely been overlooked[1].

*Lehrstuhl für Statistik, Wirtschaftswissenschaftliche Fakultät, Universität Regensburg, D-93040 Regensburg

[1]As an exception see Hamerle and Nagl (1987) who explicitly considered time invariant covariates.

In different contexts, however, the effects of covariates on the properties of the corresponding estimators have been investigated. In the framework of linear regression models it was shown (e.g. Zyskind, 1967) that the ordinary least squares estimator is also a best linear unbiased estimator if the matrix of covariates or the covariance matrix satisfies certain conditions. In a more general context, Li and Duan (1989) have shown that the slope parameters in regression analysis under link violation can consistently be estimated and are asymptotically normal if, among other conditions, conditions concerning the distribution of the covariates hold.

In the present paper we investigate the effects of block and time invariant covariates on the properties of GEE estimators. In section 2 the GEE approach is briefly outlined and in section 3 some results with respect to the properties of the GEE estimators using invariant covariates are derived. Results of a simple simulation study are presented in section 4 and concluding remarks can be found in section 5.

2 The GEE approach

Let $Y_n = (Y_{n1}, \ldots, Y_{nT})'$ denote the $(T \times 1)$ vector of responses within the nth block $(n = 1, \ldots, N)$ and $Y = (Y_1', \ldots, Y_N')'$ the $(NT \times 1)$ vector of all NT observations[2]. Observations from different blocks are assumed to be independent. Let $X_{nt} = (X_{nt1}, \ldots, X_{ntP})'$ denote the $(P \times 1)$ vector of fix covariates associated with the ntth observation $(t = 1, \ldots, T)$, X_n the $(T \times P)$ matrix of covariates associated with the nth block and X the $(NT \times P)$ matrix having full column rank associated with all NT observations. Furthermore, $\eta = (\eta_{11}, \ldots, \eta_{NT})'$, where $\eta_{nt} = X_{nt}'\beta$ and β is a $(P \times 1)$ identifiable regression parameter vector.

The 'generalized estimating equations' for the estimation of β, are given by

$$X'\mathrm{D}\Sigma^{-1}(y - \mu) = 0,$$

where $\mu = \mathrm{E}(y) = \mathrm{h}(\eta)$ and $\mathrm{D} = \partial\mu/\partial\eta = \mathrm{diag}(\partial\mu/\partial\eta_{11}, \ldots, \partial\mu/\partial\eta_{NT})$ is a diagonal matrix. $\Sigma = \phi \mathrm{A}^{1/2}(\mathrm{I}_N \otimes \mathrm{R}_\alpha)\mathrm{A}^{1/2}$ is a block diagonal matrix, where ϕ is the dispersion parameter, R_α is a $(T \times T)$ 'working' correlation matrix, fully characterized by the vector α and assumed to be identical for all blocks, $\otimes$ denotes the Kronecker product, I_N is the $(N \times N)$ identity matrix and $\mathrm{A} = \mathrm{diag}(\mathrm{V}(\mu_{11}), \ldots, \mathrm{V}(\mu_{NT}))$, where $\mathrm{V}(\mu_{nt}) = \phi^{-1}\mathrm{var}(Y_{nt})$ and $\mathrm{var}(Y_{nt})$ is the variance[3] of Y_{nt} (for details see Liang and Zeger, 1986).

Under some regularity conditions and given consistent estimators $\hat{\alpha}$ (for a critical discussion on this point see Crowder, 1995) and $\hat{\phi}$ the GEE estimator $\hat{\beta}$ is

[2]For simplicity, we assume $T_1 = \cdots = T_N = T$.

[3]The following results can easily be extended to models with $k > 2$ ordered or unordered response categories. In this case only the dimensions of the vectors and matrices change.

consistent and asymptotically normally distributed (Liang and Zeger, 1986) with covariance matrix $N^{-1}\mathrm{G}^{-1}\mathrm{W}\mathrm{G}^{-1}$, where

$$\mathrm{G} = -\lim_{N\to\infty} N^{-1}\left(X'\mathrm{D}\Sigma^{-1}\mathrm{D}X\right)_{\beta,\,\hat{\alpha}}$$

and

$$\mathrm{W} = \lim_{N\to\infty} N^{-1}\left(X'\mathrm{D}\Sigma^{-1}\mathrm{Cov}(y)\Sigma^{-1}\mathrm{D}X\right)_{\beta,\,\alpha},$$

where $\mathrm{Cov}(y) = \mathrm{diag}(\mathrm{Cov}(y_1),\ldots,\mathrm{Cov}(y_N))$ and $\mathrm{Cov}(y_n)$ is the 'true' covariance matrix of y_n.

The estimate $\hat{\beta}$ is iteratively calculated switching between a modified Fisher scoring for β and the estimation of α. Given current estimates $\hat{\alpha}_j$ and $\hat{\beta}_j$ ($j = 1,2,\ldots$), $\hat{\beta}_{j+1}$ is calculated by

$$\hat{\beta}_{j+1} = \hat{\beta}_j + \left(\left(X'\mathrm{D}\Sigma^{-1}\mathrm{D}X\right)^{-1}X'\mathrm{D}\Sigma^{-1}(y-\mu)\right)_{\hat{\beta}_j,\,\hat{\alpha}_j}.$$

3 Block and time invariant covariates

Block invariant covariates are often used to model time effects in the data matrix of a regression model or to model the effects of 'within'-factor levels realized in an experimental design with repeated measurements on this factor. As is usual, e.g. in a one-way analysis of variance model with repeated observations on this factor and appropriate restrictions on the parameters, we assume $X_n = Z\ \forall n$ and Z to be regular. In this case $X = (1_N \otimes Z)$, where $1_N = (1_1,\ldots,1_N)$ is a $(N\times 1)$ vector. Because D and A are both functions of η, not of y,

$$\frac{\partial\mu}{\partial\beta} = (1'_N \otimes Z')(\mathrm{I}_N \otimes \mathrm{D}_n),$$

where $\mathrm{D}_n = \mathrm{D}_{n'} = \mathrm{diag}(\partial\mu_{n1}/\partial\eta_{n1},\ldots,\partial\mu_{nT}/\partial\eta_{nT})$ and $\Sigma = \phi\mathrm{I}_N \otimes (\mathrm{A}_n^{1/2}\mathrm{R}_\alpha\mathrm{A}_n^{1/2})$ $\forall n,n'$. Then, G and W may be rewritten as

$$\mathrm{G} = -\lim_{N\to\infty} N^{-1}\phi^{-1}(1'_N \otimes Z')(\mathrm{I}_N \otimes \mathrm{D}_n)(\mathrm{I}_N \otimes \Sigma_n^{-1})(\mathrm{I}_N \otimes \mathrm{D}_n)(1_N \otimes Z)$$

and

$$\begin{aligned}\mathrm{W} \;=\; &\lim_{N\to\infty} N^{-1}\phi^{-2}(1'_N \otimes Z')(\mathrm{I}_N \otimes \mathrm{D}_n)(\mathrm{I}_N \otimes \Sigma_n^{-1})\mathrm{Cov}(y)\\ &(\mathrm{I}_N \otimes \Sigma_n^{-1})(\mathrm{I}_N \otimes \mathrm{D}_n)(1_N \otimes Z).\end{aligned}$$

Under the usual assumption that all necessary inverses exist, rearranging terms yields

$$\mathrm{G}^{-1}\mathrm{W}\mathrm{G}^{-1} = \lim_{N\to\infty} N^{-1}\left((1'_N \otimes (\mathrm{D}_n Z)^{-1})\mathrm{Cov}(y)(1_N \otimes (Z'\mathrm{D}_n)^{-1})\right)_\beta. \qquad (1)$$

Given current estimates $\hat{\alpha}_j$ and $\hat{\beta}_j$, $\hat{\beta}_{j+1}$ is then given by

$$\hat{\beta}_{j+1} = \hat{\beta}_j + N^{-1}\left((1'_N \otimes (\mathrm{D}_n Z)^{-1})(y-\mu)\right)_{\hat{\beta}_j}. \tag{2}$$

From (1) and (2) it can be seen that the asymptotic covariance matrix as well as the estimate $\hat{\beta}$ is independent of the 'working' correlation matrix. That is, the GEE estimator calculated under the assumption of independence is identical to an GEE estimator calculated under the assumption of any other correlation structure and has the same asymptotic — and estimated asymptotic — covariance matrix. Furthermore, it can easily be shown that in this case the asymptotic covariance matrix is identical to the asymptotic covariance matrix of 'quasi-likelihood' estimators (see McCullagh, 1983), for which McCullagh (1983) claimed — within a class of estimators for which the influence function is linear, i.e. estimators satisfying $\hat{\beta} - \beta = \mathrm{L}_\mu(Y-\mu) + o_p(N^{-1/2})$, where L_μ is a $P \times NT$ matrix of influences — to have minimum asymptotic variance.

Along the same line of arguments, the independence of the estimators and their asymptotic covariance matrices from the modelled structure of dependence in the presence of block invariant covariates as defined above, can also be shown to hold in the framework of multivariate generalized linear models using the robust covariance matrix defined by White (1982). For example, in the classical linear model it can easily be shown that the necessary and sufficient condition of Theorem 1 in Zyskind (1967) holds, and therefore the ordinary least squares estimator is also best linear unbiased.

Using covariates which are constant over T and restricting $\beta_t = \beta_{t'}\ \forall t, t'$, a similar although not as far-reaching result can be shown. In this case $X = (Z \otimes 1_T)$, where Z is a $(N \times P)$ matrix of covariates, $\partial\mu/\partial\eta = \mathrm{D}_t \otimes \mathrm{I}_T$, where $\mathrm{D}_t = \mathrm{D}_{t'} = \mathrm{diag}(\partial\mu_{1t}/\partial\eta_{1t}, \ldots, \partial\mu_{Nt}/\partial\eta_{Nt})\ \forall t, t'$ and $\Sigma = \phi(\mathrm{A}_t^{1/2} \otimes \mathrm{I}_T)(\mathrm{I}_N \otimes \mathrm{R}_\alpha)(\mathrm{A}_t^{1/2} \otimes \mathrm{I}_T)$, where $\mathrm{A}_t = \mathrm{A}_{t'} = \mathrm{diag}(\mathrm{V}(\mu_{1t}, \ldots, \mathrm{V}(\mu_{Nt}))\ \forall t, t'$.

Inserting and rearranging terms yields

$$\begin{aligned}\mathrm{G}^{-1}\mathrm{W}\mathrm{G}^{-1} \;=\; & \lim_{N\to\infty} N^{-1}\Big((((Z'\mathrm{D}\mathrm{A}^{-1}\mathrm{D}Z)^{-1}(Z'\mathrm{D}\mathrm{A}^{-1})) \otimes ((1'_T\mathrm{R}_{\hat{\alpha}}^{-1}1_T)^{-1} \\ & (1'_T\mathrm{R}_\alpha^{-1})))\mathrm{Cov}(y)(((\mathrm{A}^{-1}\mathrm{D}Z)(Z'\mathrm{D}\mathrm{A}^{-1}\mathrm{D}Z)^{-1}) \otimes \\ & ((\mathrm{R}_\alpha^{-1}1_T)(1'_T\mathrm{R}_{\hat{\alpha}}^{-1}1_T)^{-1}))\Big)_\beta \end{aligned} \tag{3}$$

and $\hat{\beta}$ may iteratively be estimated using

$$\begin{aligned}\hat{\beta}_{j+1} \;=\; & \hat{\beta}_j + N^{-1}\Big((((Z'\mathrm{D}\mathrm{A}^{-1}\mathrm{D}Z)^{-1}(Z'\mathrm{D}\mathrm{A}^{-1})) \otimes ((1'_T\mathrm{R}_{\hat{\alpha}_j}^{-1}1_T)^{-1} \\ & (1'_T\mathrm{R}_{\hat{\alpha}_j}^{-1})))(y-\mu)\Big)_{\hat{\beta}_j}. \end{aligned} \tag{4}$$

From (3) and (4) it can be seen that the properties of the GEE estimators as well as their asymptotic covariance matrix are functions only of the sums over columns and over rows and columns, respectively, of R_α and $\mathrm{R}_{\hat{\alpha}}$. For example,

it can easily be shown that the GEE estimators calculated under the assumption of an equicorrelation structure and under the assumption of independence are identical and, moreover, also their asymptotic — and estimated asymptotic — covariance matrices are identical. In this case, however, the asymptotic covariance matrix does not reduce to the asymptotic covariance matrix of quasi likelihood estimators.

The same results can again be shown to hold in the framework of multivariate generalized linear models assuming a covariance structure as in the GEE approach. This assumption is very restrictive. However, in classical linear models $\Sigma_n = \Sigma_{n'}, \forall n, n'$, and the results from above apply. Moreover, if the model is correctly specified and the 'true' correlation structure is an equicorrelation structure the ordinary least squares estimator and the generalized least squares estimator are identical and are fully efficient. Again, the condition of Theorem 1 in Zyskind (1967) can be shown to hold.

4 A Simulation Study

In order to get a hint about how the properties of the GEE estimators are affected if both types of covariates, i.e. block and time invariant covariates are present, we conducted a simulation study using the 'interactive matrix language' (IML) included in the SAS system ('statistical analysis system'), version 6 (SAS Institute Inc., 1989).

Samples were generated according to an analysis of variance model with one 'within'-factor having three ($T = 3$) levels and an additional time invariant covariate (Model I). For every of the $s = 200$ replications $N = 1000$ objects were generated. A design matrix was created having full column rank, modeling symmetric restrictions on the effects of the 'within'-factor levels. The time invariant covariate was generated as a normally distributed variate. The 'true' values of the parameters were $\beta_1 = -.3$ for the constant term, $\beta_2 = 1$ and $\beta_3 = -1$ modeling the effects of the levels of the 'within'-factor and $\beta_4 = 1$ weighting the time invariant covariate in Model I. We also simulated a 'reference' model (Model II) with a constant term ($\beta_1 = -.3$) and three free varying covariates, namely two trichotomous covariates ($\beta_2 = 1$ and $\beta_3 = -1$) with values -1, 0 and 1, generated with equal probability, and a normally distributed covariate ($\beta_4 = 1$). The error terms were generated as standard normally distributed variates with equal correlations between t and t' ($\rho_{tt'} = .8$). Using this high 'true' value and large samples, differences between the estimators are more likely to appear than with low values for $\rho_{tt'}$ and small samples (e.g. Spiess and Hamerle, 1995).

The 'observable' responses Y_{nt} were binary, where

$$Y_{nt} = \begin{cases} 1 & \text{if } Y^*_{nt} > 0, \\ 0 & \text{otherwise} \end{cases}$$

and Y^*_{nt} is the response variable of the generated latent linear model. These specifications lead to a binary probit model with an underlying equicorrelation structure.

As estimators we calculated the maximum likelihood estimator (ML estimator) for the random effects probit model (e.g. Butler and Moffit, 1982), restricting the error variance to unity, and the GEE estimators under the assumption of independence and an equicorrelation structure in the observable responses, respectively. For the ML estimator, denoted as $\hat{\theta} = (\hat{\beta}_1, \ldots, \hat{\beta}_4, \hat{\sigma})'$, where $\sigma^2 = \rho_{tt'}$, to be unbiased in this model, a necessary condition is a sufficient number of points used for the approximative evaluation of the intergrals in the log likelihood function and their derivatives. To ensure this, we calculated the ML estimates for the models over $s = 200$ replications, successively increasing the number of evaluation points by one until the results remained stable.

Table 1: *Mean (m), estimated standard deviation ($\widehat{sd}$), standard deviation of the estimates (sd) and mean of t-values (m_t) for Model I and Model II with* $N = 1000$, $T = 3$, $\rho_{tt'} = .8$, $\beta_1 = -.3$, $\beta_2 = 1$, $\beta_3 = -1$ *and* $\beta_4 = 1$ *over* $s = 200$ *replications*

m $\widehat{\text{sd}}$ sd m_t	Model I GEE_I	 GEE_E	 ML	Model II GEE_I	 GEE_E	 ML
$\hat{\beta}_1$	−.3025	−.3021	−.3021	−.3026	−.3034	−.3028
	0.0367	0.0366	0.0364	0.0386	0.0388	0.0378
	0.0384	0.0380	0.0380	0.0390	0.0383	0.0379
	−8.25	−8.26	−8.30	−7.83	−7.82	−8.01
$\hat{\beta}_2$	1.004	1.004	1.003	1.007	1.006	1.006
	0.0384	0.0383	0.0382	0.0456	0.0432	0.0418
	0.0367	0.0364	0.0363	0.0463	0.0450	0.0425
	26.15	26.20	26.28	22.13	23.34	24.06
$\hat{\beta}_3$	−1.003	−1.003	−1.003	−1.004	−1.004	−1.002
	0.0401	0.0401	0.0399	0.0454	0.0430	0.0417
	0.0368	0.0267	0.0355	0.0442	0.0431	0.0415
	−25.02	−25.05	−25.12	−22.13	−23.37	−24.03
$\hat{\beta}_4$	1.002	1.001	1.001	1.007	1.007	1.006
	0.0456	0.0455	0.0449	0.0432	0.0410	0.0400
	0.0451	0.0454	0.0445	0.0457	0.0459	0.0434
	22.04	22.07	22.28	23.36	24.61	25.15
$\hat{\sigma}$			0.8955			0.8950
			0.0167			0.0176
			0.0167			0.0195
			54.13			51.92

The GEE estimators, denoted as $\hat{\theta} = \hat{\beta} = (\hat{\beta}_1, \ldots, \hat{\beta}_4)'$ were calculated as described in Section 3, but unlike Liang and Zeger (1986) or Sharples and Breslow (1992) we estimated α starting with the Pearson correlation matrix of the residuals $\widehat{\mathrm{R}}$. Under the usual assumptions, this correlation matrix is guaranteed to be positiv definite.

The off-diagonal elements of $\widehat{\mathrm{R}}$ were then Z- transformed (Fisher, 1963) to get unbiased estimates. If all off-diagonal elements are restricted to the same value (i.e. $\alpha = \alpha$, $0 < |\alpha| < 1$ and $\alpha \neq -1/(T-1)$), the resulting correlation structure is an equicorrelation structure in the observable response variables. In this case $\hat{\alpha}$ was calculated as $\hat{\alpha} = (\exp(2\bar{z}) - 1)/(\exp(2\bar{z}) + 1)$ where $\bar{z}$ is the arithmetic mean of the Z-transformed off-diagonal elements of the matrix $\widehat{\mathrm{R}}$. The corresponding GEE estimator will be denoted GEE_E estimator. The restriction $\alpha = 0$ of course leads to an GEE estimator calculated under the assumption of independence, which will be denoted GEE_I estimator.

To compare the results, we used the following measures: (1) the arithmetic mean of the estimates over $s = 200$ replications (m), (2) the estimated standard deviation defined as $\widehat{\mathrm{sd}} = (s^{-1} \sum_{r=1}^{s} \widehat{\mathrm{var}}(\hat{\theta}_{kr}))^{1/2}$, where $\widehat{\mathrm{var}}(\hat{\theta}_{kr})$ is the estimated asymptotic variance of the kth element of $\hat{\theta}_r$ $(r = 1, \ldots, s)$, (3) the standard deviation of the estimates over the replications (sd) and (4) the arithmetic mean of the t-values over the replications (m_t).

If both block and time invariant covariates are present (Model I) the GEE_I and the GEE_E estimators are not identical and are not as efficient as the ML estimator, although the differences are very small (see Table 1). The differences between the GEE_I, GEE_E and ML estimators are larger for Model II, with the ML estimator beeing the most efficient estimator in terms of smaller $\widehat{\mathrm{sd}}$ and sd and larger values m_t followed by the GEE_E estimator.

5 Conclusions

The results from section 3 show that the properties of the GEE estimators are not independent of the kind of covariates included into the model. Therefore, when comparing GEE estimators calculated under different assumptions of correlation structures with each other or with other estimators the different effects of different kinds of covariates have to be taken into consideration.

In practical applications the results imply that if only block invariant covariates are included into the model and if one is only interested in the regression parameter vector β, the GEE estimator under the assumption of independence should be choosen. If only time invariant covariates are included into the model the calculation of the estimates as well as the asymptotic covariance matrix and their estimation are influenced only marginally by using different structures for the 'working' correlation matrix. Therefore, the efficiency gain by choosing the 'correct' vs. an 'incorrect' correlation structure can be expected to be only negli-

gible. In the case of time invariant covariates, however, the GEE estimators are not guaranteed to be as efficient as an ML estimator.

If both time and block invariant covariates as defined in section 3 above are included into the model, e.g. like in a two-way analysis of variance model with one 'between' and one 'within' factor, it is not as easy as in section 3 to derive corresponding asymptotic properties. The results of section 4 suggest that in the case of a binary probit model the differences between different GEE estimators and the ML estimator are still negligible compared to the differences of the different estimators in a model with only free varying covariates.

References

Butler, J.S. & Moffit, R. (1982). Notes and comments: A computationally efficient quadrature procedure for the one-factor multinomial probit model. *Econometrica, 50(3)*, 761–764.

Crowder, M. (1995). On the use of a working correlation matrix in using generalised linear models for repeated measures. *Biometrika, 82(2)*, 407–410.

Fisher, R.A. (1963). *Statistical methods for research workers* (13th ed). Edinburgh: Oliver and Boyd.

Fitzmaurice, G.M., Laird, N.M. & Rotnitzky, A.G. (1993). Regression models for discrete longitudinal responses (with discussion). *Statistical Science, 8(3)*, 284–309.

Hamerle, A. & Nagl, W. (1987). *Misspecification in models for discrete panel data: Applications and comparisons of some estimators* (Diskussionsbeiträge Nr. 105/s). Konstanz: Fakultät für Wirtschaftswissenschaften und Statistik.

Li, K.-C. & Duan, N. (1989). Regression analysis under link violation. *The Annals of Statistics, 17(3)*, 1009–1052.

Liang, K.-Y. & Zeger, S. (1986). Longitudinal data analysis using generalized linear models. *Biometrika, 73(1)*, 13–22.

McCullagh, P. (1983). Quasi-likelihood functions. *The Annals of Statistics, 11(1)*, 59–67.

SAS Institute Inc. (1989). *SAS/IML software: Usage and reference, version 6.* Cary, NC: SAS Institute.

Sharples, K. & Breslow, N. (1992). Regression analysis of correlated binary data: Some small sample results for the estimating equation approach. *Journal of Statistical Computation and Simulation, 42*, 1–20.

Spiess, M. & Hamerle, A. (1995). *A comparison of different methods for the estimation of regression models with correlated binary responses* (Regensburger Beiträge zur Statistik und Ökonometrie Nr. 37). Regensburg: Wirtschaftswissenschaftliche Fakultät der Universität Regensburg.

White, H. (1982). Maximun likelihood estimation of misspecified models. *Econometrica, 50(1)*, 1–25.

Zeger, S.L. & Liang, K.-Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics, 42*, 121–130.

Zyskind, G. (1967). On canonical forms, non-negative covariance matrices and best and simple least squares estimators in linear models. *The Annals of Mathematical Statistics, 38*, 1092–1109.